Introducing the New Version Biological Macromolecular Crystallization Database (BMCD4)

A new and improved version of the popular BMCD has been completed that has expanded to include over 8000 entries of crystal structures and other macromolecules. BMCD4 includes useful analytical capabilities along with the ability to enter important ancillary information, including macromolecule sequence, crystal growth conditions and physical and chemical properties of the crystals. The availability of BMCD4 is expected to enable more rapid innovation of new small molecule and biologic drugs.

M. Tung and T. D. Gallagher (Div 831)

S ince biological function is based on the structure of proteins, major efforts have been dedicated to the determination by x-ray crystallography of the structures of thousands of proteins from key organisms. These methods require high-quality crystals (diffracting to at least 3 Å), so efficient methods of crystallizing proteins are actively being sought, and have resulted in the accumulation of large amounts of crystallogenesis data. The Biological Macromolecular Crystallization Database (BMCD) first developed at NIST in 1985 has archived crystallogenesis data principally from the Protein Data Bank (PDB), now the Research Collaboratory for Structural Bioinformatics PDB. Previous versions of the BMCD drew data from the PDB by accessing the reports on a one to one basis. Recently, improved methods of accessing and importing crystallization data from the PDB have been developed, enabling the BMCD to be updated on a more automated and frequent basis. Recent research has also focused on

determining correlations between the physical and chemical properties of a protein and its tendency to form crystals yielding high resolution x-ray diffraction data for structural determination.

NIST expects the hit number on the BMCD4 website to be greater than the 1 million per year for the current version.
(http://xpdb.nist.gov:8060/BMCD4)

The BMCD has now been updated to a β 4.0 test version (BMCD4) and now includes over 8000 entries with crystallization data on proteins and other macromolecules. New entries in the BMCD4 have been expanded to include macromolecule sequence and sequence-derived information, enabling the analysis of relationships between protein properties, crystal growth conditions, and the geometric and diffraction properties of crystals. BMCD4 (http://xpdb.nist.gov:8060/BMCD4) also offers enhanced search capabilities that can retrieve specified subsets of entries for analysis. For example, utilizing the expanded content and other database features, a strong correlation was found between the distributions of crystal density and diffraction resolution.

Impact: It is expected that with the expansion of the entries in the PDB (eg. recently about 4,000 entries per year) that the BMCD4 will expand accordingly. It is anticipated that the number of hits of this new version of the BMCD would be greater than the number of hits of 1 million per year of the former version (2.0) of the BMCD.

Future Plans: It is planned to enhance new features on the text search engine in order to perform numeric range searches such as protein crystallization at a specific pH level, and to perform searches with different citation formats. It is planned to add a multidimensional backend to display statistical and summarized features of the database and query-selected subsets, rather than displaying just a list of entries.